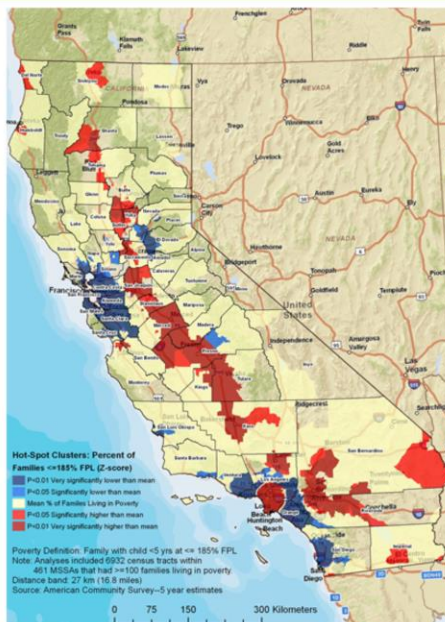


# Analyzing patterns in your data

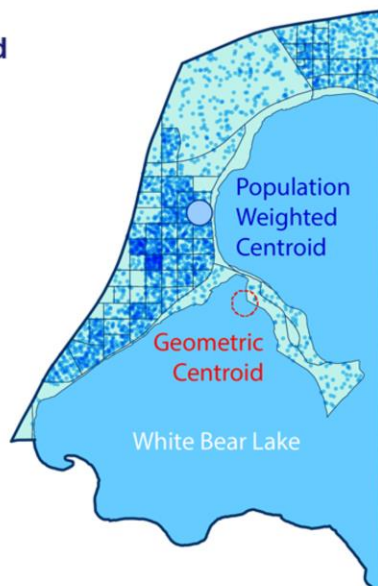
Using GIS Training to Address Blood  
Pressure Medication Adherence



- Conceptualize where your population lives
- Understand scale issue in spatial studies
- Get to know spatial autocorrelation
- Examine and interpret hotspot analysis



- **Population Weighted Centroid (population center) –**
  - center point of a region's population
- **Simplicity is good**

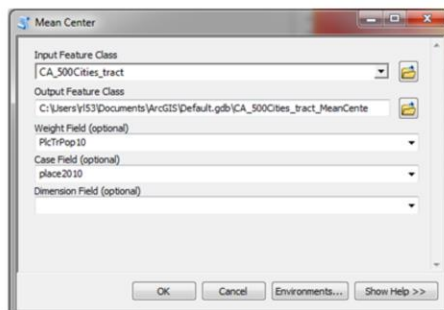


Points are easier to work with. The relationship to other feature can be simplified.

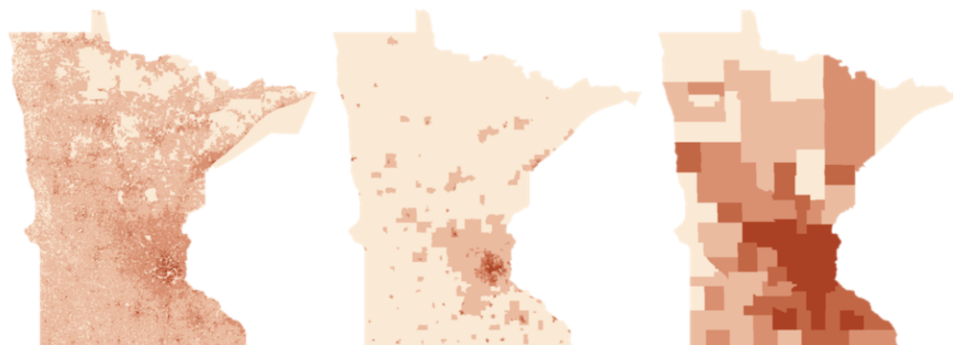
- Use population in sub-unit as weight
- Make sure you have the identifier for the polygons
- Other variables can be used as weight

$$(X, Y) = \frac{\sum_{i=1}^n p_i (x_i, y_i)}{\sum_{i=1}^n p_i}$$

\* where  $(X, Y)$  is the location for PWC,  
 $(x_i, y_i)$  is population location in the region,  
and  $p_i$  is the population for point  $i$



Population weighted centroid (population center) data is available at census website: <https://www.census.gov/geo/reference/centersofpop.html>



**Block**

**Census Tract**

**County**

- Different scales produce different maps
- Data will exhibit different patterns at different scales

Maps showing here are population density for MN at different scale. Maps are displaying in quintile classification across all maps.

- **Spatial data is rarely independent**
- **Tobler's First Law of Geography**
- **Neighborhood should be taken into consideration**



Tobler's First law of Geography: Everything is related to everything else, but near things are more related than distant things

- Use Getis-Ord  $G^*$  index

$$G_i^* = \frac{\sum_{j=1}^n \omega_{ij} x_j}{\sum_{j=1}^n x_j}$$

- Compares observed  $G^*$  with expected  $G^*$

$$Z_i^* = \frac{G_i^* - \overline{G_i^*}}{\sqrt{\text{Var}(G_i^*)}}$$

- Shows statistical significance

*\* where  $x_j$  is the local measure (prevalence),  
 $\omega_{ij}$  is spatial weight matrix.*

- Hot spots in red and cold spots in blue

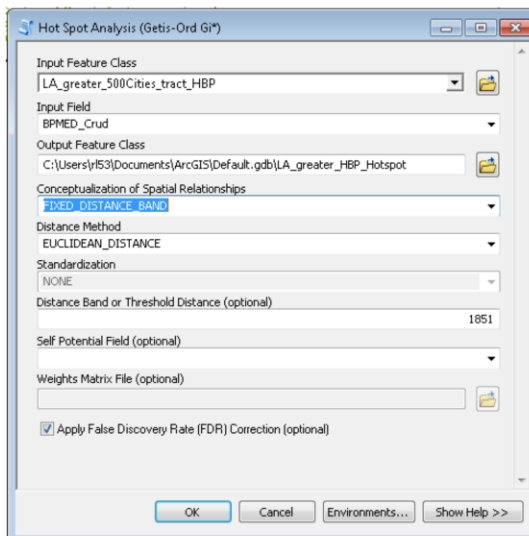


Hotspot analysis uses vectors (not rasters)

Use Getis-Ord  $G^*$  index (taking neighborhood value into account)

Density can tell you where clusters in your data exist, but not if your clusters are statistically significant

- Hotspot analysis is a diagnostic tool
- Spatial relationship settings is crucial
  - Distanced based: Fixed Distance Band, and etc.
  - Polygon Contiguity: edge/ edge corner
  - Get Spatial Weights From File
- Multi-test correction



Distanced based - Etc. : Inverse Distance, Inverse Distance Squared, Zone of Indifference

Euclidean distance weighted – Closer features are weighed more heavily than features that are further away

**Fixed Distance Band** – Every feature within a fixed distance is included, every feature outside that distance is excluded

**Zone of Indifference** – Combination of Inverse Distance and Fixed Distance Band

Polygon Contiguity – Only features that share a border are included

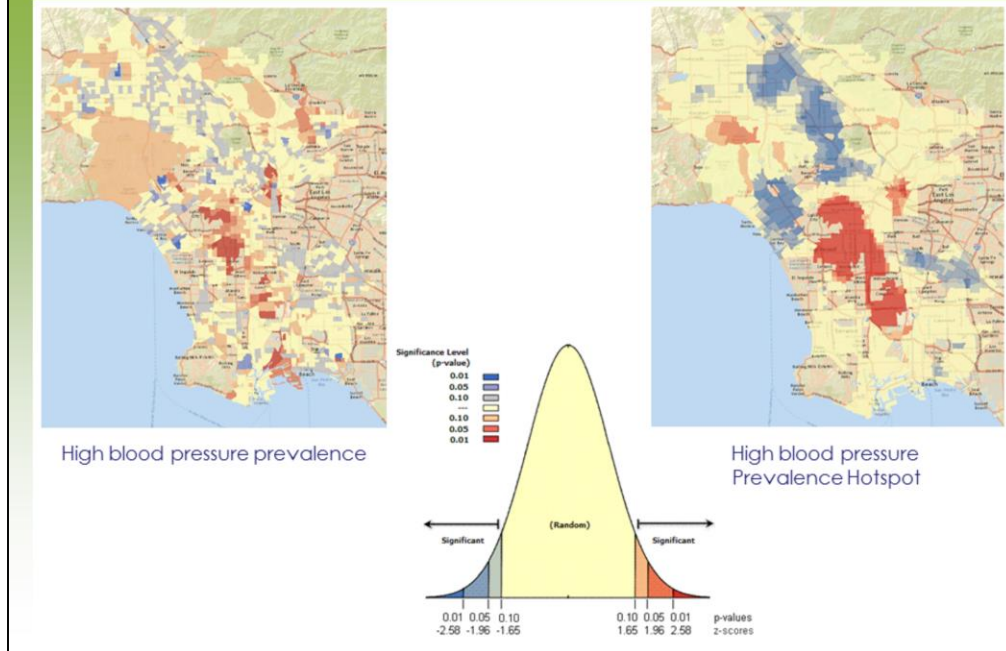


## Define Study Scale

- **Data must be projected!**
- **Distance unit needs to match the projection**
- **“Knowledge” is the key**
- **Useful information is the goal**
- **Driving distance could be used to estimate community**



In the exercise, we want to look at hot spots/ cold spots for high blood pressure prevalence in greater LA region. We used average of the driving distance to nearest 5 pharmacies from census population weighted centroids. Specifically speaking, 3529 meters



A high Z score and small P value for a feature indicates a significant hot spot.

A low negative Z score and small P value indicates a significant cold spot. The higher (or lower) the Z score, the more intense the clustering.

A Z score near zero means no statistically significant spatial clustering.

Table

LA\_greater\_500Cities\_tract\_1

BPMED_Crud	Shape_Length	Shape_Area	GiZScore Fixed 1851	GiPValue Fixed 1851	Gi_Bin Fixed 1851_FDR
71	4533.927505	1050465.083991	1.330017	0.183513	0
65.4	2728.779012	451632.328309	1.194478	0.232291	0
70.4	2742.95001	436640.356434	1.234646	0.216962	0
71.3	5330.488501	1285547.479277	1.187154	0.235167	0
71	5802.202296	1713239.466485	1.057894	0.290104	0
69.4	4353.563057	1091611.442068	0.139884	0.888752	0
66.3	4563.43631	827728.574154	0.974731	0.329694	0
66.8	3962.201182	951408.084695	1.425619	0.153978	0
67.4	3511.182224	521033.1522	2.107365	0.035086	0
73.1	3598.572421	803827.854204	2.371197	0.017731	1
74.2	4845.34673	1049642.672061	2.068716	0.038573	0
68.3	4272.264813	934623.405216	1.88012	0.060092	0
76.1	7491.019607	2187506.376554	1.976332	0.048117	0
69.5	3683.598281	655706.313976	1.967176	0.049163	0
69	3381.987394	623012.936758	1.924939	0.054237	0
70.9	4265.521528	765503.815726	2.535946	0.011214	1
70.9	4521.406272	862068.363237	2.396788	0.01654	1
75.4	3965.940505	822043.376808	2.107862	0.035043	0
71.3	6014.36457	1833958.561435	1.493716	0.13525	0
71.9	5413.679254	888097.430013	2.162431	0.030585	0

1 (0 out of 1511 Selected)

LA\_greater\_500Cities\_tract\_1

Resulting z-scores and p values are stored in the attribute table of your output.

The “Bin” attribute field contains the categories you can use to display hotspots (considers the z-scores and p values to create the categories).

You don’t want to compare either Z score and P value, Gi\_bin is the place to look at